

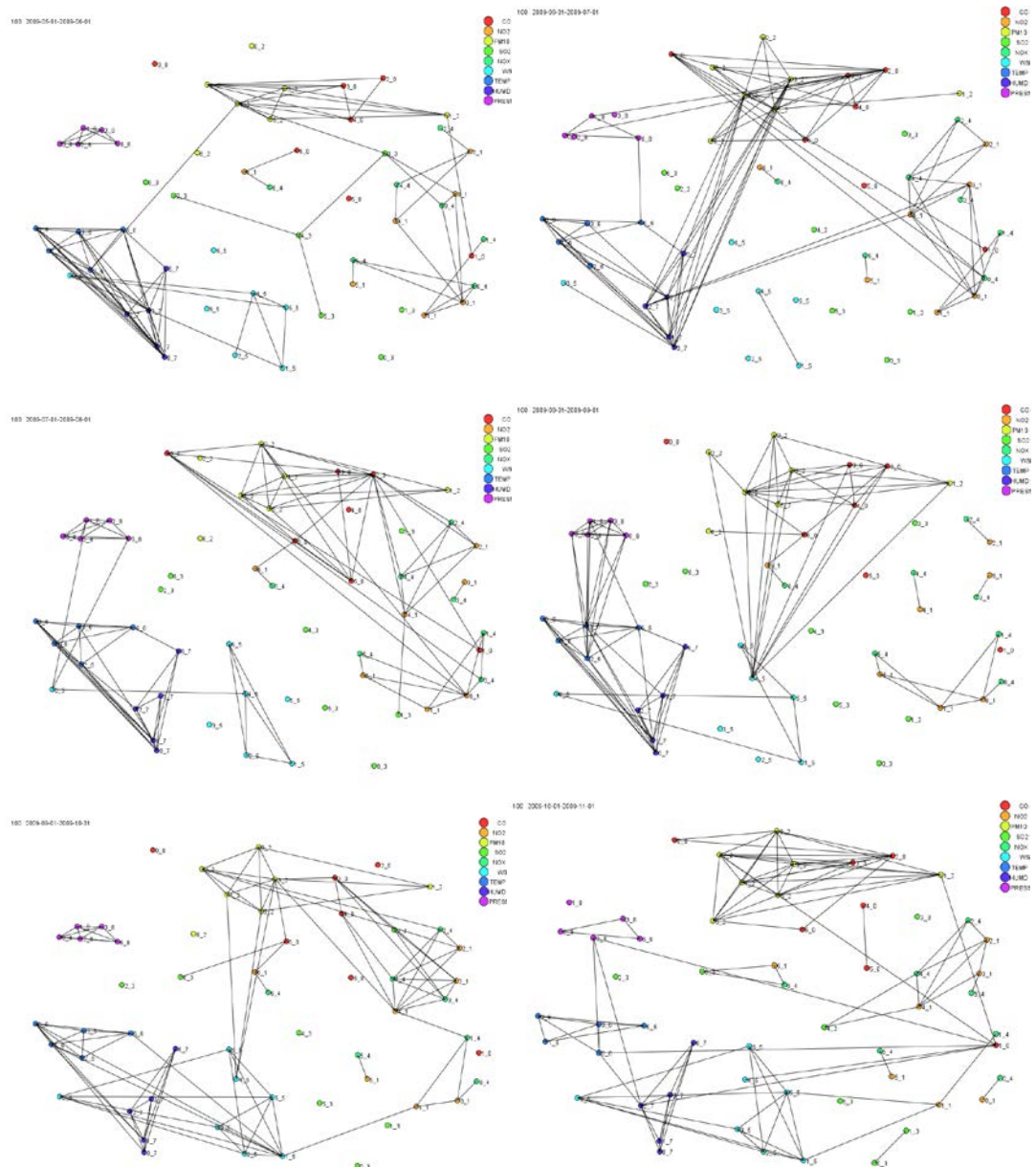
Weekly Report (2014.05.05-2014.05.11)

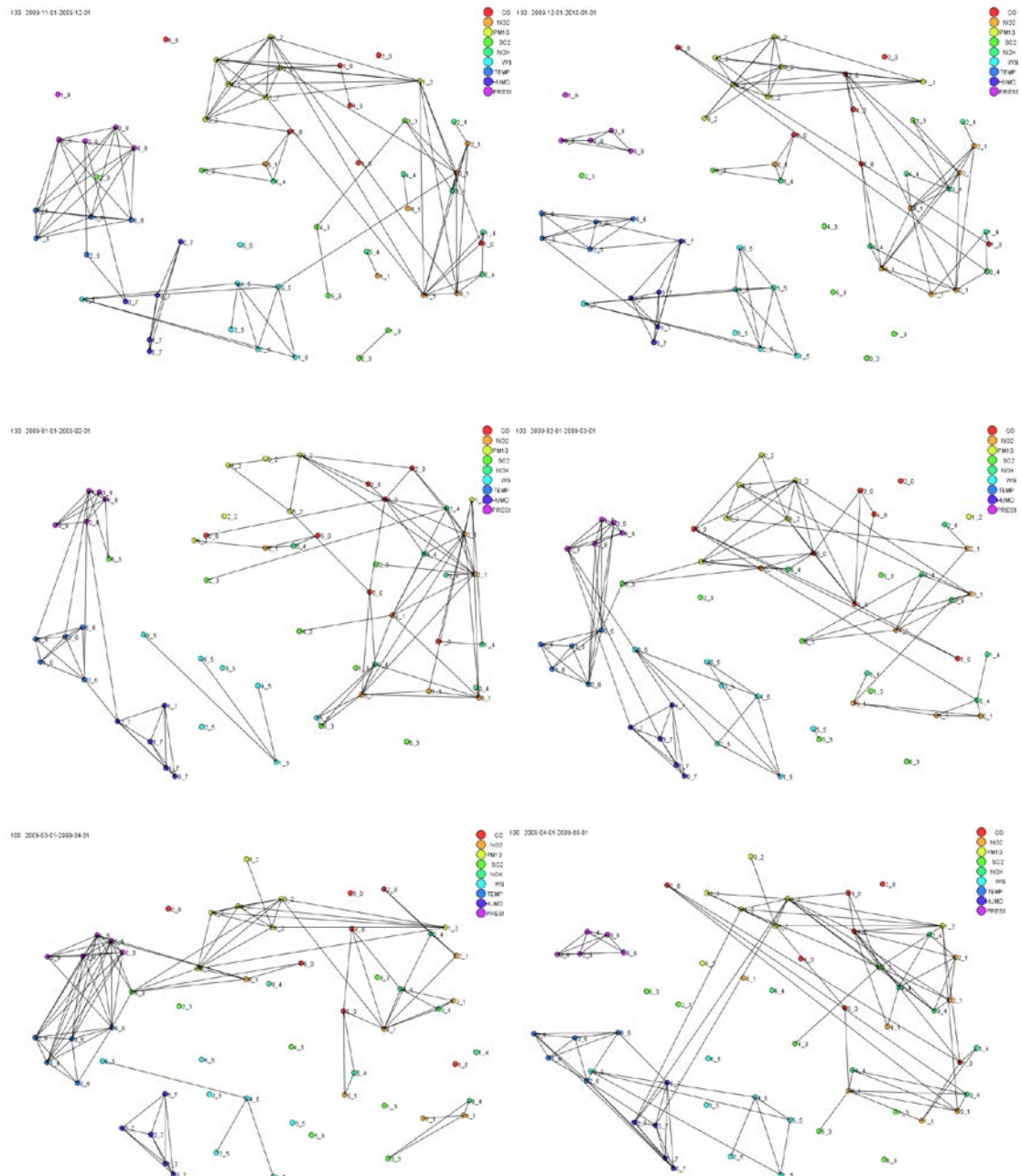
Guo Fangzhou

1. TCP Tree to TCP Graph

This week we decided to draw a graph of all variables (57 variables in total) in the data first. The MDS was applied to draw the graph and we took the mutual information as dissimilarities. Pairs of nodes that have the top 100 smallest mutual information are linked as edges. Nodes that present different attributes are filled with different color.

We took one month as time step and used data in 2009, therefore we finally got 12 graphs, as shown below:





From the graphs we can see that there are two groups of nodes, one is the nodes that have attributes of WS, TEMP, HUMD or PRESS, the other is the nodes that have attributes of CO, NO2, PM10, SO2 or NOX. Besides, there are some communities in the graphs. The links between the communities are time varying and the communities themselves are also time varying.

2. Mobile Data

Heatmap was implemented and the problem of different cells has different refreshing speed was solved by removing repeated time adjacent records which belong to the same user. The graph below is the heatmap of Lucheng district at 8:00 AM. It is very clear that there are two intersections that were very crowded at the time.



We thought that the movement from one cell to another cell can be treated as a vector. All the movements could compose a vector field but after we tried we found that the visualization of this “vector field” is bad.

3. Future Data

I tried Cointegration Test but it doesn't seem to be useful.